



Examining the relationship of textbooks and labs on student achievement in eighth grade science

Don Jones, Ed.D.

Anacita Sugalan, Ed.D.

Marie-Anne Mundy, Ph.D.

LaVonne Fedynich, Ed.D.

Texas A&M University – Kingsville

Abstract

One of the most important objectives of teachers, parents, school administrators, and students is to improve student scores on standardized tests, such as the State of Texas Assessment for Academic Readiness (STAAR) in eighth-grade science. This quasi-experimental study examined the science achievement scores between schools that used different textbooks and labs when delivering instruction. This study utilized a quantitative approach, using archival data and survey design. Analysis of covariance (ANCOVA) and multiple regression were used to analyze the data while controlling STAAR eighth-grade reading scores to reveal significant differences between classes. The sample and population for this study were predominantly eighth-grade Hispanic students in South Texas.

Analysis of covariance showed that classes that used labs with high hand-on experiences with greater direct student participation received higher science scores on state assessments. Additionally, reading scores were significantly related to science scores. Multiple regression findings indicated that textbooks and labs were significant predictors of student achievement on the STAAR eighth-grade science class result in South Texas for Spring 2015. The findings of this study may serve as a catalyst for improving student achievement in science through changes in textbook adoption and doing labs in science. The result suggests the need to research further to investigate other contributing factors of student achievement.

Keywords: Science Instruction, Science Curriculum, Textbook Selection, NCLB, Assessment, Science Achievement

Introduction

Low student achievement in science, as shown by the State of Texas Assessment for Academic Readiness (STAAR) results, has concerned many school leaders. Many articles have been written about the challenges of science education. Students in the United States dropped significantly in international standing for science achievement from elementary to high school (Martin et. al, 2011). In general, findings have shown that students in the United States have had alarmingly poor performances in science content and process knowledge on standardized assessments when compared to European and Asian countries (Gonzalez et. al, 2009; Organization for Economic Co-operation and Development, 2009).

A need exists to address this concern because students lose interest in science as they get older, which may affect the outcome of the test scores (U.S. Department of Education, 1992), and students are not completing school as proficient in science (O'Neil, 1992). Children begin school with curiosity about the

world around them and have an interest in discovering things, but over the years, that enthusiasm for science may be lost, and as they progress to high school, students take the minimum number of science courses required for graduation (U.S. Department of Education, 1992).

Texas Education Agency (2014) reports that statewide passing rates for the STAAR mathematics and reading in grades 3-8 remained stable for the past three consecutive years. Scores in fifth-grade science remained the same at 73%, while eighth-grade decreased by 4%. For Texas school districts, the lack of progress on STAAR is consistent, regardless of the student demographics.

Educators in Texas are seeking solutions to improve scores in science. Discussions on U.S. education policy are focused on how the quality and effectiveness of science and mathematics curriculum and resources can be improved. Reports from the National Research Council (NRC, 2012) and the National Governors Association (NGA, 2011) suggested possible strategies to increase the participation and preparation of students in science, technology, engineering, and mathematics (STEM) fields. Some researchers have pointed out that the problem lies within the curriculum because school districts are not explicit about what the curriculum should be or how it should be implemented (Saphier, Specia, & Gower, 2008). Nearly all are in agreement that something is wrong with the state's testing system as students' test scores have not improved (Johnson, Johnson, & Johnson, 2012). Others blame teachers for being less prepared (O'Neil, 1992) and suggest the use of hands-on activities or labs to improve science achievement (Guzman & Bartlett, 2012).

The Association for Supervision and Curriculum Development (2013) mentioned that in 1983, the National Commission on Excellence Education (NCEE) examined the quality of teaching and learning in elementary and secondary education and reported in *A Nation at Risk* that the educational system in the United States was declining. The report presented the failure of its schools to produce students able to compete in the global economy. To rectify this problem, the NCEE suggested increasing high school requirements in core subject areas, including science, and updating textbooks to include more rigorous and challenging material. The curriculum in high school was added with four years of English, three years of math, three years of science, three years of social studies, and half of a year of computer education.

In addition, the No Child Left Behind (NCLB) Act (2001) was passed to strengthen and redefine American education by addressing four basic tenets: accountability for results, emphasis on doing science research, parental options, and more local control and flexibility. The NCLB (2001) required that, within a decade, all students would be proficient on state academic assessments (Simpson, La Cava, & Graner, 2004). However, results in the Program for International Student Assessment (PISA) in 2009 placed the United States rank 22 in science proficiency behind Norway, Japan, Hungary, and Slovenia (Center on Educational Policy, 2008). The United States' diminishing world standing in science should be a concern because this trend will continue if mathematics, reading, and language arts continue to be emphasized over science (Perry & McConney, 2010).

With science tested at the fifth grade, eighth grade, and high school levels (Metz, 2011), students are not learning science foundations because schools spend a disproportionate amount of instructional time on mathematics, reading, and language arts instead. Research shows subjects like science, which are not included in the assessment programs, are most likely to be ignored, resulting in students having less than favorable attitudes toward them (Barmby, Jones, & Kind, 2007).

The misconceptions, anxieties, and lack of interest in science manifest from a lack of needed science exposure early on in students' education (Mallow et al., 2010). There are fewer people pursuing science-related fields, and the low number of women and minorities entering the sciences and engineering

programs is a major concern in the United States as well (Wagner, 2010). Therefore, better prepared professionals and encouraging more students to pursue careers in science could help the United States retain its economic competitiveness in the world's market economy (National Research Council, 2012). This issue points to the need for a stronger focus from school leaders involved in decisions related to curricular and textbook evaluation and adoption.

Purpose of the Study

The objective of this quasi-experimental study was to determine whether utilizing different textbooks and labs affects student achievement in eighth-grade science in four different districts in south Texas in order to ultimately inform and improve science achievement in the region. Archived data from spring 2015 eighth-grade science STAAR test results were used to measure student success. This study identified the textbook that impacted the most on eighth-grade science STAAR test and identified that the use of labs as a pedagogical approach in teaching science affected students' overall achievement. The STAAR test results gave the researcher a clear understanding of factors that possibly impacted student achievement.

Research Questions

In quantitative studies, research questions and hypotheses are used to shape and focus the purpose of the study (Creswell, 2014). Research questions aided the researchers in determining the relationship between variables. For the purpose of this study, labs were classified as low, medium, or high in relation to the degree to which students were involved in "hands-on" lab activities. The questions addressed in this study were as follows:

RQ1. Is there a difference among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson Interactive and other resources on student achievement as measured by the 2015 STAAR eighth-grade class science results while controlling for the 2015 STAAR eighth-grade class reading score in south Texas?

RQ2. Is there a difference among schools that utilized low labs, medium labs, and high labs on student achievement as measured by the 2015 STAAR eighth-grade science class results while controlling for the 2015 STAAR eighth-grade reading score in south Texas?

RQ3. Is there a relationship between the criterion variable of student achievement as measured by the 2015 STAAR eighth-grade science class results and the predictor variables of classes' utilization of textbooks and labs in south Texas?

Null Hypotheses

The following hypotheses were used in this study to determine the effect of the textbook and labs used in student achievement:

H₀₁: There is no significant difference among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson Interactive and other resources on student achievement as measured by the 2015 eighth-grade class science results while controlling for the 2015 STAAR eighth-grade class reading score in south Texas.

H₀₂: There is no significant difference among schools that utilize low labs, medium labs, and high labs on student achievement as measured by the 2015 eighth-grade science class results while controlling for the 2015 STAAR eighth-grade reading score in south Texas.

H₀₃: There is no significant relationship between the criterion variable of student achievement as measured by the 2015 STAAR eighth-grade science class results and the predictor variables of classes' utilization of textbooks and labs in south Texas.

Theoretical Framework

Understanding how students learn and factors that influence science achievement may result in improved students' science STAAR scores. The theoretical framework of this study was based on the premise that individuals learn through interaction with each other and their environment. This viewpoint, that students are active thinkers who construct their own understanding from interactions with phenomena, the environment, and other individuals, is based on the theory of constructivism (Piaget, 1970).

Jean Piaget popularized the theory of constructivism wherein the variables influencing how students construct their knowledge are taken into account. According to the constructivism theory of learning, students learn by constructing meaning from what they experience (Solomon, 2003). The constructivist theory incorporates the cognitive and social aspects of a student's environment with a student's specific interest to create a more holistic approach to learning. Constructivist theory is rooted in the idea that in order for learning to be long lasting, it must be relevant and meaningful (Gentry & Springer, 2002). As education takes on a more personal approach, more students are given opportunities to gain their educational experience in a constructivist manner. Hence, using constructivist theory as the premise, this study focused on performance differences in student learning as it related to the degree to which lab experiences were high hands-on versus those that had lower hands-on experiences.

Research Design and Approach

A quantitative quasi-experimental study research design (posttest only) was used to determine the impact of Science Fusion, Stemscope, IScience, Pearson Interactive, Pearson Interactive and other resources, and the use of labs on student achievement as measured by the 2015 STAAR eighth-grade science class results for students in south Texas. Non-equivalent groups were compared using ex post facto data posttest only, but utilizing a covariate to equalize the groups instead of randomization. Student test data on the 2015 STAAR eighth-grade science class results from the four different districts using different textbooks, as well as labs, were collected and analyzed to see if they predicted student achievement. The independent variables were represented by the textbook and labs used. The dependent variable was represented by student achievement. Student achievement data were taken from the 2015 spring administration of the STAAR eighth grade science.

Setting and Sample

This study utilized eighth-grade science students from four different school districts in south Texas. The study involved 25 schools, and each school has at least two to four science classes. The four districts have similar demographics, which include enrollment, racial make-up, and socio-economic distribution that made the districts appropriate for comparison. The sample comprised a total of 71 intact groups of classes of eighth-grade students: approximately 6,945 students during the academic years 2014-2015. There were 341 students that were excluded from the sample because they did not have reading scores. The sample population was a convenience sample since the researcher used naturally formed groups provided by the school district (Creswell, 2014), and they were chosen based on their availability and the characteristics that the researcher wanted to investigate (Gall et al., 2003). Moreover, the researcher was a science teacher in one of the districts in south Texas, and the population consisted of eighth-grade students who were enrolled in science.

The sample size was determined based on the rule of thumb for determining the number of subjects required for statistical analyses (Van Voorhis & Morgan, 2007). Having a larger sample size increases the power and it also represents the characteristic of the populations from which they are derived (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Marcoulides, 1993). The independent variables, which were the books, has four cells (textbook A, textbook B, textbook C, and textbook D), and the labs have three cells (low labs = 0-15%, medium labs = 16-30%, and high labs = 31-40%). There were four districts composed of twenty-five schools with a total of 72 science classes. Classes have 71 cells (A, B, C, D, E, F, G, H). To detect differences between, or among groups, 25-30 participants per cell should be the rule of thumb and this lead to about 80% power (Cohen, 1988). Therefore, the required sample size was approximately 45-60 classes for an alpha level of .05, a confidence interval of 95%, and a power of about 0.8 (Creswell, 2014), which can be seen in Table 1.

Table 1
Classes Distribution for the School Year 2014-2015 (N=72)

District	# of Schools	# of Classes	% of Total Sum
A	8	24	33%
B	7	18	25%
C	6	18	25%
D	4	12	17%
Total	25	72	100%

Archival Data

For the first two research questions, the researcher obtained archival data regarding students' STAAR scores in science and reading in the 2015 STAAR eighth-grade science and reading test from the principals of each school after an approval from the superintendent of the district to conduct the study was obtained. Before the approval of the superintendent, the researcher had to seek approval from Texas A & M University - Kingsville (TAMUK), who oversaw and made sure that all federal, institutional, and ethical guidelines were followed. For some schools that were not complete with the data, the researcher had to contact the Public Information Office of each district and their individual procedures and requirements for data access were followed before data was released. The data was provided in an Excel format and sent to the researcher electronically. In addition, eighth-grade science teachers from each school were also contacted by the researcher through emails, phone calls, and personal visits to inquire about the names of the textbooks and the percentages of labs utilized in science.

Validity and Reliability

To ensure the validity of the STAAR assessment, committees made up of educators from different districts across the state convened, and they served as advisors for each grade level and content area. Committees of teachers, staff members from the Texas Education Agency (TEA) and test development specialists were formed to identify the Texas Essentials Knowledge and Skills (TEKS) expectations that were important to assess. This group of educators collaborated and provided input to the content of the testing items and the state curriculum standards. The questions that were chosen to be included on the test were extensively reviewed by TEA, its testing contractor, and other Texas educators throughout the state of Texas to make sure that they aligned to the TEKS (TEA, 2013). This alignment of content and assessment provided validity evidence that STAAR assessment measures the student achievement.

The reliability of the STAAR test is based on the Kuder-Richardson Formula-20 (KR-20), which measures internal consistency. This Formula-20 is considered by educational and psychological specialists as one of

the best instruments to measure reliability (Borg & Gall, 1989). Formula-20 is mostly used for tests that use multiple-choice items and to see if they produce the same results over a population of testing subjects. The STAAR assessment scored a reliability coefficient ranging from .87 to .90 (TEA, 2013). Thus, the STAAR assessment is reliable because the coefficient is high.

Results

Descriptive Statistics for Archived Data

The descriptive statistics for the archived data were provided for students regarding the variables in this study, such as textbooks and labs, and science scores for Spring 2015 (see Table 2). The total dataset ($N = 72$) was included in the analysis. The textbook was coded as 0 = Science Fusion, 1 = Stemscope, 2 = IScience, and 3 = Pearson Interactive and 4 = Pearson and other resources. Labs were coded as Low = 0-15%, Medium = 16-30%, and High = 31-40%.

Table 2

Frequency Description for Student's Class Scores and Textbook Used ($N = 72$)

Textbook	Mean	Standard Deviation	N
0 (Science Fusion)	61.10	19.89	24
1 (Stemscope)	66.06	18.55	18
2 (IScience)	60.50	18.50	18
3 (Pearson Interactive)	68.27	16.81	6
4 (Pearson + other)	62.95	17.30	6
Total	62.86	18.92	72

The average grade for the overall eighth-grade classes was ($M = 62.86$, $SD = 18.92$). Classes that utilized Pearson Interactive obtained significantly higher science class results ($M = 68.27$, $SD = 16.81$) than classes that utilized any other textbook. Classes that utilized Stemscope obtained significantly higher science class results ($M = 66.06$, $SD = 18.54$) than classes that utilized Science Fusion ($M = 61.10$, $SD = 19.89$), and classes that utilized IScience ($M = 60.50$, $SD = 18.50$). Classes that utilized Pearson and other resources scored significantly better than classes that utilized IScience. There were 24 classes that utilized Science Fusion, 18 classes utilized Stemscope, 18 classes utilized IScience, six classes utilized Pearson Interactive and there are also six classes that utilized Pearson and other resources (See Table 2).

Information is provided in Table 3 regarding utilization of labs. The researcher separated the labs into three categories: low labs = 0-15%, medium labs = 16-30%, and high labs = 31-40%. The distribution for lab utilization was: two classes (3%) utilized low labs, 43 classes (59%) utilized medium labs, and 27 classes (38%) utilized high labs. The overall average science scores for all classes were $M = 62.86$ and $SD = 18.76$. Classes that utilized high labs obtained significantly the highest class science scores ($M = 66.05$, $SD = 18.76$). Classes that utilized medium labs obtained science scores that were slightly higher ($M = 60.87$, $SD = 18.87$) than classes that utilized low labs ($M = 60.41$, $SD = 18.63$).

Table 3

Frequency Description for Student's Class Scores and Labs Used ($N = 72$)

Labs	Mean	Standard Deviation	N	% of Total Sum
Low (0-15%)	60.41	18.63	2	3%
Medium (16-30%)	60.87	18.75	43	59%
High (31-40%)	66.05	18.76	27	38%
Total	62.86	18.92	72	100%

Inferential Statistics

Research question 1.

The following question guided this quantitative study:

Is there a difference among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson Interactive and other resources on student achievement as measured by the 2015 STAAR eighth-grade class science results while controlling for STAAR 2015 eighth-grade class reading scores in South Texas?

The following null hypothesis was quantitatively tested:

H_{01} : There is no significant difference among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson Interactive and other science resources on student achievement as measured by the 2015 STAAR eighth-grade class science results while controlling for the 2015 STAAR eighth-grade class reading scores in South Texas.

A one-way analysis of covariance (ANCOVA) was conducted for this study to determine the effect of textbooks on student scores. The independent variable (textbook) included five levels: Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson and other resources. The dependent variable (student achievement) was measured by the STAAR eighth-grade class science results, and the covariate was the STAAR eighth-grade class reading score. The ANCOVA was inappropriate because the test of homogeneity of slopes was significant. Therefore, the three levels representing low, medium, and high levels were selected: one standard deviation below the mean, the mean, and one standard deviation above the mean and the covariate. Then the ANCOVA was run. The interactions were statistically significant: $F(4, 7093) = 217.36, p = .00, \eta^2 = .11$. The strength of the relationship of the interaction between the textbook and reading was moderate as assessed by partial Eta-squared. There was a significant difference among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson Interactive and other science resources on student achievement as measured by the 2015 STAAR eighth-grade class science results while controlling for their 2015 STAAR eighth-grade class reading scores in South Texas ($F(4, 7093) = 221.39, p = .00, \eta^2 = .11$). The strength of the relationship between textbook and student achievement was moderate as assessed by partial Eta-squared. The textbook utilized accounted for the 11% of the variance of achievement scores. Information is provided in Table 4.

Table 4

Test of Between- Subjects Effects- Textbook

Source	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Reading	1	2723.14	.000	.28
Textbook	4, 7093	221.39	.000	.11
Text * Reading	4	217.36	.000	.11
Error	7093			

Note: $p < .05$

In conclusion, the null hypothesis was rejected. On looking at the three different values, only one textbook, the Pearson Interactive, obtained significantly higher scores in science at all 3 reading levels.

Research question 2.

The following question guided this quantitative study:

Is there a difference among classes that utilized low labs, medium labs, and high labs on student achievement as measured by the STAAR eighth grade science class results while controlling for the 2015 STAAR eighth grade reading scores in south Texas?

The following hypothesis was quantitatively tested:

H₀₂: There is no significant difference among classes that utilize low labs, medium labs, and high labs on student achievement as measured by the 2015 eighth-grade science class results while controlling for the 2015 STAAR eighth-grade reading scores in South Texas.

A one-way analysis of covariance (ANCOVA) was conducted for this study to determine the effect of labs on student achievement. The independent variable (labs) included three levels: low, medium, and high labs. The dependent variable (student achievement) was measured by the STAAR eighth grade science class results. The ANCOVA was statistically significant: $F(2, 7097) = 34.52, p = .000, \eta^2 = .01$ (see Table 5). The strength of the relationship between the lab used and student achievement was low as assessed by partial Eta squared. The labs used accounted for 1% of the variance of achievement score.

Table 5

Test of Between-Subjects Effects- Labs

Source	<i>df</i>	<i>F</i>	<i>p</i>	η^2
Reading	1	1073.31	.000	.13
Labs	2	50.14	.000	.01
Labs * Reading	2	34.52	.000	.01
Error	7097			

In conclusion, the null hypothesis was rejected. For the labs at reading level 1, high labs scored significantly higher than medium and low labs. At medium reading levels, with adjustments made for multiple comparisons, only high lab usage scored significantly better than medium labs. At high reading levels, with adjustments made for multiple comparisons, no significant differences were found.

Research question 3.

The following question guided this quantitative study:

Is there a relationship between the criterion variable of student achievement as measured by the 2015 STAAR eighth-grade science class results and the predictor variables of class utilization of textbooks and labs in South Texas?

The following null hypothesis was quantitatively tested:

H₀₃: There is no significant relationship between the criterion variable of student achievement as measured by the 2015 STAAR eighth-grade science class results and the predictor variables of class utilization of textbooks and labs in South Texas.

A multiple regression analysis was conducted to determine how well the variables textbook and labs predicted student achievement as measured by the science class results. Assumptions of linearity, normally distributed errors, and uncorrelated errors were checked and met. The linear combination of textbook and labs significantly predict student achievement ($F(2, 7102) = 63.58, p < .001$), with both variables significantly contributing to the prediction. The adjusted *R* squared value was 0.017. This indicates that 1.7% of the scores in science can be predicted by labs and textbook. According to Cohen's (1988) classification, this is a very small effect size. The beta weights presented in table 4 (previously presented) suggests that high use of labs contributed to predicting good scores in science and textbooks

contribute to a lesser degree. In conclusion, the null hypothesis was rejected: textbooks and labs were significant predictors of student achievement ($p < .001$) on the STAAR eighth-grade science class result in south Texas for spring 2015, but the effect size was very small (See tables 6 and 7).

Table 6

Means, Standard Deviations, and Intercorrelations for Science Achievement and Predictor Variables ($n = 72$)

Variables	Mean	Standard Deviation	1	2
Science Achievement	62.87	18.92	.034	.127
Predictor variable				
1. Text	1.39	1.23		-.023
2. Labs	2.35	0.55		

Table 7

Simultaneous Multiple Regression Analysis Summary for Textbook and Labs in Predicting Science Achievement ($n = 72$)

Variables	<i>B</i>	<i>SEB</i>	β
Text	.561	.18	.04
Labs	4.42	.41	.13
Constant	51.71	1.01	

Descriptive Statistics for the Survey

The second section of this study involved a survey for eighth-grade science teachers in 2014-2015. The researchers obtained permission from the principals of the 25 schools in the four districts to contact science teachers through emails informing them about the study. The total number of teachers was 72. Two questions were asked: the name of the textbook they were using, and the percentage of labs used, which were categorized into low labs = 0-15%, medium labs = 16-30%, and high labs = 31-40%. The names of the textbooks were given a numerical value of 0 = Science Fusion, 1 = Stemscope, 2 = IScience, 3 = Pearson Interactive, and 4 = Pearson Interactive and other resources.

Of the 72 teachers, six teachers responded. Another email was sent to the principals and teachers asking for assistance in responding to the questions and assuring them of the confidentiality of the information. In addition, the researcher asked permission from principals to schedule a visit to the respective schools at a convenient time to meet the teachers personally during conference periods, and the principals agreed. The researcher started school visits during the second week of January. There were at least five schools that were visited on each scheduled day. At each visit, a paper with the two questions on it and a pencil were provided to the teachers, and it took them at least two minutes to answer the questions. By the last week of February, 25 schools were visited, and 72 teacher responses were recorded. The results showed that the number of teachers that used medium labs ($n = 43$) was higher compared to teachers that used high labs ($n = 27$). The number of teachers that used low labs ($n = 2$) was low compared to the teachers that used medium and high labs (see Table 8).

Table 8
Frequency Description of Teacher's Responses on Utilization of Labs (n=72)

Labs	Number of Teachers	Percent
Low	2	3%
Medium	43	59%
High	27	37%
Total	72	100%

Responses on the use of textbooks indicated that out of 72 teachers, 24 teachers (34%) employed Science Fusion, which was a considerably higher number compared to teachers who used Stemscope, which was 18 (25%). There were 18 teachers (25%) who availed of IScience. Finally, there were six (8%) teachers who utilized Pearson Interactive, and there were also six (8%) teachers who have resorted to Pearson Interactive and other resources as shown in Table 9.

Table 9
Frequency Description of Teacher's Responses on Utilization of Textbooks (n =72)

Book	Number of Teachers	Percent
Science Fusion	24	34%
Stemscope	18	25%
IScience	18	25%
Pearson Interactive	6	8%
Pearson Interactive + other	6	8%
Total	72	100%

Summary of the Findings

This study addressed three research questions and the corresponding hypotheses. Research Hypothesis 1 tested the difference among classes that utilized textbooks on student achievement as measured by the 2015 STAAR eighth-grade class science results. The difference was found to be statistically significant. Research Hypothesis 2 tested the difference among schools that used low, medium, and high labs on student achievement as measured by the eighth-grade class science results. The difference was found to be statistically significant. Research Hypothesis 3 tested the relationship between the criterion variable of student achievement as measured by the 2015 STAAR eighth-grade science class results and the predictor variables of class utilization of textbooks and labs. The relationship was found to be statistically significant. Textbook and labs predicted student achievement. For the first two hypotheses, ANCOVA was used, while multiple regression was employed for the third hypothesis.

Conclusions and Recommendations

Research Hypothesis 1 tested the significant differences among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson and other resources on student achievement as measured by the 2015 STAAR eight-grade class science results while controlling for STAAR 2015 eighth-grade class reading scores. A total of 72 classes were examined in 25 schools in four districts in south Texas, comparing the science class results of classes that utilized textbooks and labs. According to descriptive statistics, the achievement score as measured by the STAAR eight-grade science class result showed that there was a significant difference among classes that utilized Science Fusion, Stemscope, IScience, Pearson Interactive, and Pearson Interactive and other science resources on student achievement as measured by the 2015 STAAR eighth-grade class science results while controlling for the 2015 STAAR eighth-grade class reading scores in south Texas. The strength of the relationship between

the textbook and student achievement was moderate. Looking at the three different values, only one textbook, the Pearson Interactive, obtained significantly higher scores in science at all three reading levels. The findings of this research support the studies of Abd-El-Khalick et al. (2017), NRC (2006), Chiappetta, Sethna, and Fillman (1991), Stern and Ahlgren (2002), and Shiland (1997), indicating the influence of textbooks on student achievement. The result of this study showed that textbooks had a relationship with student achievement, and as suggested by Ulkerick (2015), it is important to examine the role of textbooks in teaching. Teachers in this study relied on textbooks for lesson planning, which supported the study of Weiss (1993) and Radcliffe, Caverly, Peterson, and Emmons (2004). Because of the role that textbooks play in education, a thorough evaluation of textbooks is needed to ensure that textbooks are adopted and aligned with science learning goals.

Research Hypothesis 2 tested the differences among classes that utilized low, medium, and high labs on student achievement as measured by the 2015 eighth-grade science class results. According to descriptive statistics, it appeared that classes that used high labs had a significantly higher score ($M = 66.05$), followed by classes that utilized medium labs ($M = 60.87$), and classes that employed low labs obtained the lowest science class results ($M = 60.41$). The finding of this study supports the research of Patke (2013), Bohr (2014), Guzman and Bartlett (2012), NAEP (2011), Duschl (2008), Strand-Cary and Klahr (2008), Supalo (2010), Ruby (2001), Stohr-Hunt (1996), Doran and Tamir (1992), Blosser (1990), Mullis and Jenkins (1988), Shymansky et al., (1983), Boghai (1979), Grozier (1969), and Godomsky (1971), who pointed out the direct relationship of labs to student achievement.

According to descriptive statistics, the results of this study showed that for the labs at reading level 1, high labs scored significantly higher than medium and low labs. At medium reading levels, with adjustments made for multiple comparisons, only high lab usage scored significantly better than medium labs, which supported the findings of Stohr-Hunt (1996), Ruby (2001), and Vogt (n.d), who found out that the frequency of labs was related to student achievement. At high reading levels, with adjustments made for multiple comparisons, no significant differences were found.

On the other hand, the findings of this research were inconsistent with the research of Comber and Keeves (1973), Anderson et al. (1989), La Pointe et al. (1992), Miller (2014), Keegan (2003), Gao (2014), Roehrig and Garrow (2007), Wolf and Fraser (2007), and Hofstein and Lunetta (1982, 2003), who explained that labs have no significant relationship with student scores.

Results of the study also showed significant relationships between the dependent variable (science scores) and the controlled variable (reading scores). There was a strong correlation (0.59) between the science and reading scores, which supported the study of Imam et al., (2014), Chege (2012), O'Reilly and McNamara (2007), Grabe and Stoller (2002), Bender et al., (2008), Corcoran and Mamalakis (2009), indicating the positive relationship between reading scores and science scores.

Research Hypothesis 3 tested the relationship between the criterion variable of student achievement as measured by the STAAR eighth-grade science class results. Multiple regression analysis showed that 17% of the scores can be predicted by labs and textbooks. This low effect size supports the research of Puttick et al. (2015) and Hofstein and Lunetta (1982, 2003), who explained the factors that impacts student achievement. The lack of information on how labs were taught, and how effective they were, needs to be examined (Hofstein & Lunetta, 2003). In addition, Harrell (2010) emphasized one of the most important factors that affect science scores, which is the knowledge of the teacher in implementing the curriculum.

In conclusion, the results of this study indicated that classes that used Pearson Interactive had the highest science class result. Further research should closely examine textbook contents to determine the factors that contributed to the difference in scores. As to the effect of labs on student achievement, the result of this study showed that classes that resorted to high labs had higher science class results and that textbooks and labs were significant predictors of student achievement. Therefore, the use of labs in science classrooms must be encouraged if better results are to be achieved. Administrators and science leaders must provide the resources to support the use of labs, which includes providing training or staff development to help familiarize teachers with labs needed for their lessons, and monetary resources to purchase supplies for labs or hands-on activities.

References

- Abd-El-Khalick, F., Myers, J. Y., Summers, R., Brunner, J., Waight, N., Wahbeh, N., ... Belarmino, J. (2017). A longitudinal analysis of the extent and manner of representations of nature of science in U.S. high school biology and physics textbooks. *Journal of Research in Science Teaching*, 54, 82-120. doi:10.1002/tea.21339
- Anderson, L., Ryan, D., & Shapiro, B. (1989). *The IEA classroom environment study: International studies in education achievement*, (Vol. 4). New York, NY: Pergamon Press.
- Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M., Drake, L. (2010). *The Condition of Education 2010* (NCES 2010-028). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Barmby, P., Jones, K., & Kind, P. M. (2007). Developing attitudes toward science measures. *International Journal of Science Education*, 29(2), 871-893. doi:1.1080/09500690600909091
- Bohr, T. (2014). *Teachers' Perspectives on Online Virtual Labs versus Hands-On Labs in High School Science*. (Doctoral dissertation). Available from ProQuest. (UMI No. 3615309).
- Center on Educational Policy. (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects*. Birmingham: Heldref Publications.
- Chege, E. W. (2012). *Reading comprehension and its relationship with academic performance among standard eight pupils in rural Machakos* (Doctoral dissertation, Kenyatta University). Retrieved from <http://ir-library.ku.ac.ke/handle/123456789/3722>
- Chiappetta, E. L., Sethna, G. H., & Fillman, D. A. (1991). A quantitative analysis of high school chemistry textbooks for scientific literacy themes and expository learning aids. *Journal of Research in Science Teaching*, 30, 787-797.
- Chiappetta, E. L., Sethna, G. H., & Fillman, D. A. (1993). Do middle school life science textbooks provide a balance of scientific literary themes? *Journal of Research in Science Teaching*, 30, 787-797.
- Comber, L., & Keeves, J. (1973). *Science education in 19 countries: An empirical study*. New York: Halstead Press.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: Sage Publications, Inc.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: Wiley.

- Duschl, R. A. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32, 268-291.
- Gall, M. D., Gall, J. P., & Org, W. R. (2003). *Educational research: An Introduction* (7th ed.). Boston, MA: Allyn & Bacon.
- Gentry, M., & Springer, P. M. (2002). Secondary student perceptions of their class activities regarding meaningfulness challenge, choice, and appeal: An initial validation study. *The Journal of Secondary Gifted Education*, 13(4), 192-204.
- Gonzales, P., Williams, T., Jocelyn, W., Roey, S., Kastberg, D., & Brenwald, S. (2009). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context*. National Center for Education Statistics. Retrieved from <https://www.nces.ed.gov/pubs2009/2009001.pdf>
- Guzman, K., & Bartlett, J. (2012). Using simple manipulatives to improve student comprehension of a complex biological process. *Biochemistry and Molecular Education*, 40(5), 320-327.
- Imam, O., Mastura, M., Jamil, H., & Ismail, Z. (2014). Reading comprehension skills and performance in science among high school students in Philippines. *Asia Pacific Journal of Educators and Education*, 29, 81-94.
- Hofstein, A., & Lunetta, V. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of Educational Research*, 52(2), 201-217.
- Hofstein, A., & Lunetta, V. (2003). The laboratory in science education: Foundations for the 21st century. *Wiley Periodicals Incorporated. Science Education*, 88, 28-54. doi:10.1002/sce.10106
- Johnson, B., & Christenson, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches*. (3rd ed.). Los Angeles, CA: Sage Publications.
- Johnson, W. L., Johnson, A. M., & Johnson, J. (2012). *Predicting student success on the Texas Chemistry STARR test: A logistic regression*. Retrieved from <http://files.eric.gov/fulltext/ED534647>
- Mallow, J., Kastrup, H., Bryant, F. B., Hislop, N., Shefner, R., & Udo, M. (2010). Science anxiety, science attitudes, and gender: Interviews from a binational study. *Journal of Science Education and Technology*, 19(4), 356-369. doi: 10.1007/s10956-010-9205-z
- Marcoulides, G. A. (1993). Maximizing power in generalizability studies under budget constraints. *Journal of Educational Statistics*, 18(2), 197-206.
- Martin, M., Mullis, I., Foy, P., & Stanco, G. (2011). TIMMS 2011 International Results in Science. TIMMS & PIRLS, (Ch. 2), 10.
- Metz, K. E. (2011). Young children can be sophisticated scientists. *Phi Delta Kappan*, 92(8), 68-71.
- National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. Washington DC: US Department of Education.
- National Research Council. (2006). *America's lab report: Investigations in high school science*. S. R. Singer, M. L. Hilton, and H. A. Schweingruber (Eds.). Washington, DC: The National Academies Press. <https://doi.org/10.17226/11311>.

- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The national Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13165#toc
- National Governors Association. (2011). *Building a science, technology, engineering, and math education agenda: Update of state actions*. NGA Center for Best Practice. Washington, DC: National Governors Association.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- O'Neil, J. (1992). *Science education: Schools pushed to broaden access, overhaul practice*. Curriculum update. Association for Supervision and Curriculum Development: Alexandria, VA.
- Patke, U. (2013). *Inquiry-based laboratory investigations and student performance on standardized tests in biology*. (Doctoral dissertation). Available from ProQuest LLC. (UMI No. 3600291)
- Perry, L. B., & McConney, A. (2010). Does the SES of a school matter? An examination of social economic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137-1162.
- Piaget, J. (1970). *Logic and psychology*. New York: Basic Books.
- Puttick, G., Drayton, B., & Cohen, E. (2015). A study of the literature on lab-based instruction in biology. *The American Biology Teacher*, 77, 23-18.
- Radcliffe, R., Caverly, D., Peterson, C., & Emmons, M. (2004). Improving text reading in a middle school science classroom. *Reading Improvement*, 4, 145-156.
- Saphier, J., Specia, M., & Gower, R. (2008). *The skillful teacher: Building your teaching skills*. Acton, MA: Research for Better Teaching, Inc.
- Shiland, T. W. (1997). Quantum mechanics and conceptual change in high school chemistry textbooks. *Journal of Research in Science Teaching*, 34, 535-545.
- Simpson, R. L., La Cava, P. G., & Graner, P. S. (2004). The No Child Left Behind Act: Challenges and implications for educators. *Intervention in School and Clinic*, 40(2), 67-75.
- Solomon, P. (2003). *The curriculum bridge: From standards to actual classroom practice*. Thousand Oaks, CA; Sage.
- Stern, L., & Ahlgren, A. (2002). Analysis of student' assessments in middle school curriculum materials: Aiming precisely at benchmarks. *Journal of Research in Science Teaching*, 39, 889-910.
- Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, 23(4), 488-511.
- Texas Education Agency. (2013). *2013 Accountability Rating*. Retrieved from <http://tea.texas.gov/student.assessment/staar/>
- Texas Education Agency. (2014a). *Preliminary statewide STAAR grades 3-8 passing rates relatively stable for third consecutive year*. Retrieved from http://www.esc12.net/UserFile/Server_2954298/File
- Ulkerick, S. (2015). Using textbooks for meaningful learning in science. *National Association for Research in Science Teaching*. Retrieved from <https://www.narst.org/publications/research/textbook2>
- U.S. Department of Education. (1992). *Helping your child learn science*. Retrieved from <https://www.docu-cu-archive.com-Helping-Your-Children-Learn-Science-pdf>

- Van Voorhis, C. R. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43-50.
- Wagner, T. (2010). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need-and what we can do about it*. New York, NY: Basic Books.
- Weiss, I. R. (1993). Science teachers rely on the textbook. *What research says to the science teacher, volume seven: The science technology, society, movement*. Washington, DC: National Teachers Association.

About the Authors

Don Jones (don.jones@tamuk.edu) is an associate professor of educational leadership at Texas A&M University – Kingsville. He teaches in the doctoral program for educational leadership and coordinates the superintendent certification program. He is a former superintendent from Texas.

Marie-Anne Mundy (Marie-nne.Mundy@tamuk.edu) is an associate professor of educational leadership at Texas A&M University – Kingsville. She teaches in the doctoral program in educational leadership and is the previous program coordinator. She has expertise in both quantitative and qualitative methodology and formerly served as the research coordinator for the doctoral program at Walden University.

LaVonne Fedynich (lavonne.fedynich@tamuk.edu) is a full professor of educational administration at Texas A&M University – Kingsville. She is the coordinator of the master's program in educational administration and principal certification. In addition, she serves on numerous doctoral dissertation committees.

Anacita Sugalan (anacita.sugalan@psjaisd.us) is a science teacher and curriculum writer for the Pharr-San Juan-Alamo ISD in the Texas Rio Grande Valley. She earned her doctorate in educational leadership from Texas A&M University – Kingsville.